ANALISIS REGRESI

Dasar dan Penerapannya dengan R

Robert Kurniawan
Budi Yuniarto

ANALISIS REGRESI

Dasar dan Penerapannya dengan

Robert Kurniawan
Budi Yuniarto

ANALISIS REGRESI:
Dasar dan Penerapannya dengan R
Edisi Pertama

Perpustakaan Nasional: Katalog Dalam Terbitan (KDT)
ISBN : 978-602-422-034-1
ISBN (E) : 978-602-422-387-8
13.5 x 20.5 cm
x, 218 hlm
Cetakan ke-1, September 2016

Kencana. 2016.0695

Penulis
Robert Kurniawan
Budi Yuniarto

Desain Sampul
Suwito

Penata Letak
Endang Wahyudin

Percetakan
PT Kharisma Putra Utama

Penerbit
K E N C A N A
(Divisi dari Prenadamedia Group)
Jl. Tambra Raya No. 23 Rawamangun - Jakarta 13220
Telp: (021) 478-64657 Faks: (021) 475-4134
e-mail: pmg@prenadamedia.com
www.prenadamedia.com
INDONESIA

KATA PENGANTAR

R merupakan salah satu program yang lebih banyak dikembangkan untuk bidang statistik dan *R* adalah program yang bersifat *open source* yang sangat mudah digunakan bagi pemula. Di lain sisi, banyak peneliti yang terkendala dari sisi lisensi jika ingin menggunakan program yang terkait dengan bidang statistik. Nah, buku ini cocok untuk bagi Anda yang ingin menggunakan program *R* untuk aplikasi statistik khususnya untuk analisis regresi dan korelasi.

Seperti yang kita ketahui, regresi dan korelasi adalah sebuah alat analisis statistik yang sering digunakan oleh para peneliti, atau mahasiswa S-1 ataupun mahasiswa pasca-sarjana untuk menyusun penelitiannya. Dalam buku ini, dikupas tuntas terkait dengan analisis regresi baik dari sisi teori dan penerapannya dengan berbagai contoh yang mungkin bisa digunakan sebagai acuan untuk belajar. Aplikasi *R* yang digunakan pun sangat mudah untuk diaplikasikan sebagai praktik dari teori yang dituliskan dalam buku ini.

Buku ini membahas semua tentang analisis regresi dan beberapa analisis korelasi, yang terbagi menjadi 9 bab. Kami harapkan dengan buku ini, banyak kalangan akademisi dan umum bisa memahami statistik khususnya untuk analisis regresi secara runtut dan mengerti bagaimana cara menginterpretasikan untuk output yang dihasilkan dari program *R* tersebut. Silakan Anda bermain-main dengan buku ini untuk pemahaman Anda tentang analisis regresi dan korelasi.

Akhirnya, kami menyadari bahwa buku ini masih mengandung banyak kelemahan dan kekurangan. Oleh sebab itu, kami sangat berterima kasih apabila para pembaca bersedia memberikan kritik dan saran untuk menyempurnakan buku ini pada edisi berikutnya. Pada kesempatan ini pula kami mengucapkan banyak terima kasih kepada berbagai pihak yang telah mendorong ditulisnya buku ini. Tidak lupa pula kami mengucapkan banyak terima kasih kepada penerbit Prenadamedia Group yang telah bersedia menerbitkan buku ini. Semoga bermanfaat.

Jakarta, November 2015

Penulis

DAFTAR ISI

1

PENGENALAN R

1.1 TENTANG R

R adalah suatu sistem untuk melakukan analisis statistik dan grafik yang pertama kali dibuat oleh Ross Ihaka dan Robert Gentleman dari Statistics Department of The University of Auckland, New Zealand, pada tahun 1992. *R* merupakan suatu *software* sekaligus bahasa pemrograman, dan merupakan salah satu "dialek" dari bahasa S. Secara umum, sintaks dari bahasa *R* adalah ekuivalen dengan paket statistik S plus, sehingga sebagian besar keperluan analisis statistika, dan pemrograman dengan *R* adalah hampir identik dengan perintah yang dikenal di S plus.

R sampai saat ini merupakan hasil kolaborasi seluruh kontributor yang berasal dari seluruh dunia, karena sifatnya yang *open source*, sehingga semua orang bisa ikut berkontribusi dengan menyumbangkan kode, perbaikan *bugs*, ataupun membuat dokumentasi tentang *R*.

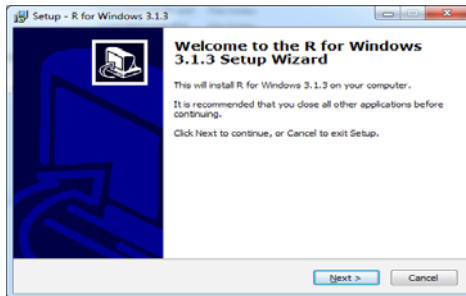
Sistem *R* bersifat multiplatform, bisa berjalan di segala jenis platform UNI*x* (misal, Linux), Windows, ataupun Mac OS. *File* instalasi *binary* tersedia untuk sistem operasi-sistem operasi tersebut, dan bisa diunduh secara gratis di http://cran.r-project.org/.

1.2 INSTALASI R DI WINDOWS

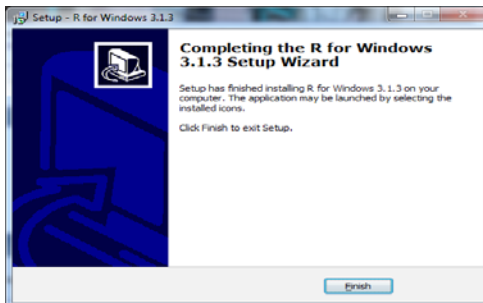
Untuk meng-*install R* pertama kali, diperlukan *file* instalasi *binary* sebagai *base distribution* yang bisa diunduh pada alamat di atas. Untuk sistem operasi Windows, maka

pilihlah *file* instalasi *binary* versi *R* untuk Windows, yaitu *R-3.1.3-win.exe*.

Setelah mengunduh *file* instalasi *binary* tersebut di atas, berikut ini langkah-langkah instalasi R:

1. Klik ganda/*double click* pada *file R-3.1.3-win.exe* yang telah diunduh, sehingga akan muncul tampilan sebagai berikut:

Gambar 1.1 Kotak Dialog Awal Instalasi R

Klik *Next*, kemudian ikuti langkah berikutnya sesuai petunjuk yang diberikan. Selama proses instalasi, Anda bisa tetap menggunakan pilihan-pilihan *default* yang diberikan.

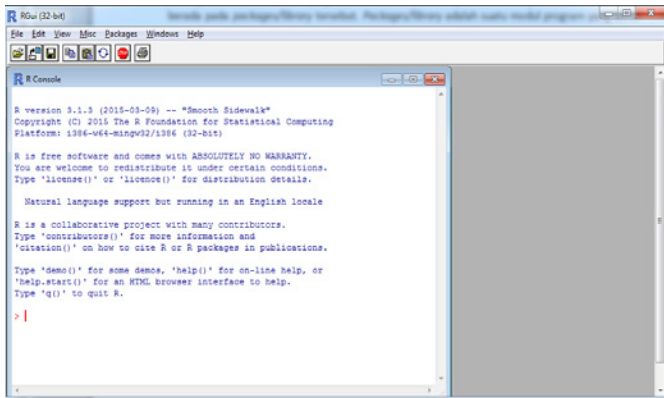
2. Setelah selesai akan muncul tampilan berikut:

Gambar 1.2 Kotak Dialog Akhir dari Proses Instalasi R

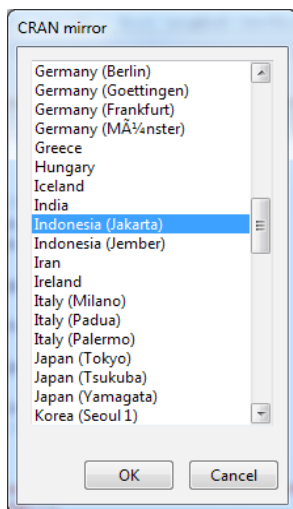
Klik *Finish* untuk mengakhiri proses instalasi.

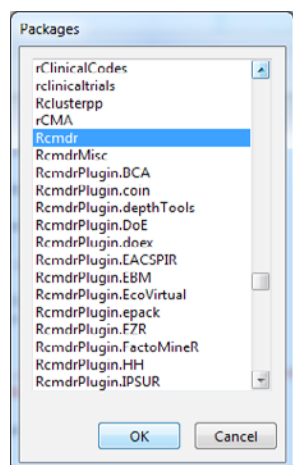
1.3 INSTALASI PACKAGES R-COMMANDER

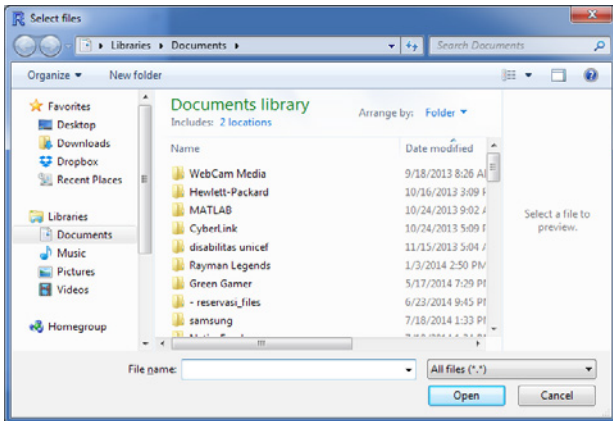
Setelah instalasi *file binary* sebagai *base* dari *R*, kita juga perlu meng-*install contributed packages/library* tambahan yang diperlukan, karena sebagian besar fungsi dan kemampuan *R* berada pada *packages* tersebut. *Packages* adalah suatu modul program yang berisi kumpulan perintah atau fungsi untuk melakukan suatu analisis tertentu menggunakan *R*. Dalam pembahasan buku ini, kita akan memerlukan *package R-commander* yang menyajikan GUI sederhana untuk keperluan beberapa analisis statistika tertentu dalam *R*.

Proses instalasi *library* tersebut ikuti langkah berikut:

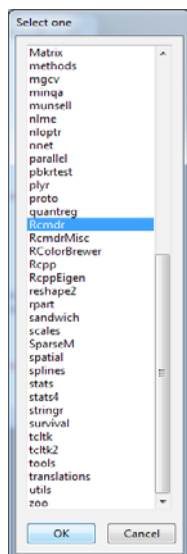
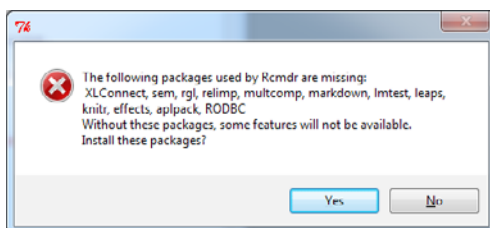
1. Jalankan program *R*, sehingga terbuka tampilan R seperti berikut:

Gambar 1.3 Tampilan R

2. Untuk instalasi secara *online*, pilih menu **Packages → Install Package(s)**. Jika kita belum mengatur pengaturan *mirror website* CRAN yang akan kita gunakan, maka akan muncul tampilan untuk memilih *mirror*. Jika pada satu *mirror* kita gagal melakukan instalasi, silakan ganti dengan *mirror* lainnya.

Gambar 1.4 Tampilan untuk Memilih *Mirror*

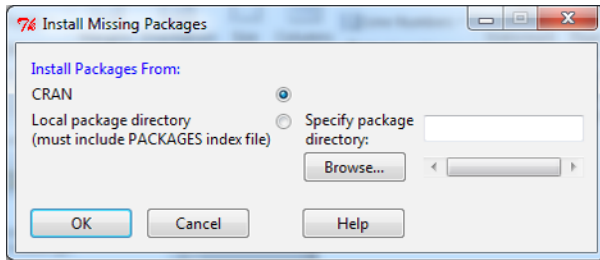
3. Setelah itu akan muncul pilihan *packages* yang ingin kita *install*. Dalam hal ini, carilah *packages Rcmdr*, kemudian klik *OK*.

Gambar 1.5 Tampilan untuk Memilih *Packages*

4. Alternatif lain adalah kita *download zip file*-nya terlebih dahulu kemudian kita lakukan instalasi secara *offline* dengan memilih menu **Packages → Install Package(s) from local zip files.** Kemudian akan terbuka jendela dialog memilih *zip files* dimaksud, sebagai berikut:

Gambar 1.6 Jendela Dialog Memilih *Local Zip File*

Namun hal ini tidak direkomendasikan, karena *dependencies* (*package* lain yang dibutuhkan untuk menjalankan suatu *package*) dari *package* Rcmdr sangat banyak.

5. Untuk mengecek apakah sudah terinstal dengan benar, maka pilih menu **Packages → Load Package(s)**. Jika sudah terinstal, maka pada tampilan seperti berikut ini akan terdapat pilihan Rcmdr. Pilih Rcmdr kemudian klik *OK*. (lihat Gambar 1.7)
6. Jika setelah klik *OK* muncul pesan seperti pada Gambar 1.8, maka berarti bahwa *packages* lain yang dibutuhkan untuk menjalankan *R-commander*. Klik *Yes* untuk meng-*install packages-packages* tersebut.

Gambar 1.7 Tampilan *Load Package*(s)

Gambar 1.8 Pesan Kesalahan Saat *Load Packages*

7. Setelah klik *Yes*, akan muncul kotak dialog untuk memilih dari mana instalasi akan dilakukan. Karena kita belum memiliki *zip file* dari *packages* tersebut, maka kita pilih CRAN untuk meng-*install* secara *online* (pastikan Anda memiliki koneksi internet). (lihat Gambar 1.9)

8. Setelah selesai, maka *R* dengan *package R-commander* sudah ter-*install* siap digunakan.

Gambar 1.9 *Install Missing Packages*

1.4 MANAJEMEN DATA DALAM R

1.4.1 Jenis-jenis Objek

R bekerja menggunakan objek, di mana semua entitas *R* termasuk fungsi dan struktur data merupakan objek dalam *R*. Objek-objek tersebut mempunyai karakteristik nama, konten, serta atribut yang menspesifikasikan tipe data yang direpresentasikan oleh objek tersebut.

Semua objek memiliki dua atribut intrinsik, yaitu *mode* dan *length*. Mode menunjukkan tipe dasar dari elemen-elemen objek dan *length* menunjukkan banyaknya elemen yang ada dalam objek. Terdapat empat *mode* utama, yaitu *numeric*, *character*, *complex*, dan *logical*. Dalam R terdapat *mode* la-innya, tetapi mode tersebut tidak merepresentasikan data.

Berikut ini adalah tabel yang menyajikan *overview* untuk tipe-tipe objek yang merepresentasikan data:

Objek	Mode	Apakah satu objek bisa memiliki lebih dari satu mode?
vector	*numeric, character, complex* atau *logical*	Tidak
factor	*numeric* atau *character*	Tidak
array	*numeric, character, complex* atau *logical*	Tidak

lanjutan ...

matrix	*numeric, character, complex* atau *logical*	Tidak
data frame	*numeric, character, complex* atau *logical*	Ya
time series	*numeric, character, complex* atau *logical*	Tidak
list	*numeric, character, complex, logical, function, expression, ...*	Ya

Vector

Vector merupakan suatu himpunan elemen (bilangan, *character* atau *string*, *logical value*) satu dimensi dan merupakan representasi dari satu variabel yang umum dikenal. *Vector* merupakan objek paling dasar yang dikenal dalam **R**. Pada data *vector* berlaku menggunakan mode tunggal, sehingga jika dua data atau lebih yang memiliki mode berbeda maka *R* akan mengubah data ke mode yang lebih umum.

Untuk mengetahui mode suatu objek *vector* dapat dilakukan dengan menggunakan fungsi atau *command* **mode**. Jumlah atau panjang data yang bertipe *vector* dapat diketahui dengan memanfaatkan fungsi **length.**

Factor

Factor merupakan representasi dari variabel kategorik. *Factor* tidak hanya mengandung nilai yang terkait variabel kategorinya, tetapi juga kemungkinan setiap level yang ada dalam variabel tersebut, meskipun level tersebut tidak muncul dalam data. Kita juga bisa menambahkan *label* pada *value level*-nya.

Untuk membuat suatu *factor* digunakan fungsi ***factor*** berikut:

factor(x, levels = sort(unique(x), na.last = TRUE), labels = levels, exclude = NA, ordered = is.ordered(x))

Contoh:

> factor(1:3)
[1] 1 2 3
Levels: 1 2 3
> factor(1:3, levels = 1:5)
[1] 1 2 3
Levels: 1 2 3 4 5
> factor(1:3, labels = c("A", "B", "C"))
[1] A B C
Levels: A B C
> factor(1:5, exclude = 4)
[1] 1 2 3 NA 5
Levels: 1 2 3 5

Matrix

Array adalah suatu tabel yang tersusun dari vektor dengan *k* dimensi, sedangkan *matrix* adalah kasus khusus dari *array* di mana *k* = 2 (*array* 2 dimensi). Untuk membuat suatu *matrix* digunakan fungsi ***matrix*** berikut:

matrix(data = NA, nrow = 1, ncol = 1, byrow = FALSE, dimnames = NULL)

Contoh:

> matrix(data = 5, nr = 2, nc = 2)
[,1] [,2]
[1,] 5 5
[2,] 5 5
> matrix(1:6, 2, 3)
[,1] [,2] [,3]
[1,] 1 3 5
[2,] 2 4 6
> matrix(1:6, 2, 3, byrow = TRUE)
[,1] [,2] [,3]

```
[1,] 1 2 3
[2,] 4 5 6
```

Data Frame

Data frame adalah suatu tabel yang dibentuk dari satu atau lebih vektor dan/atau dengan *length* yang sama tetapi dengan *mode* yang berbeda. *Data frame* bisa dibuat secara implisit menggunakan fungsi *read.table* atau kita juga bisa menggunakan fungsi *data.frame*.

Contoh:

```
> x <- 1:4; n <- 10; M <- c(10, 35); y <- 2:4
> data.frame(x, n)
x n
1 1 10
2 2 10
3 3 10
4 4 10
> data.frame(x, M)
x M
1 1 10
2 2 35
3 3 10
4 4 35
```

Time-Series

Time-series adalah objek yang berisi data runtun waktu. Fungsi ***ts*** digunakan untuk membuat objek *time series* dari suatu vektor (untuk *single time-series*) ataupun dari matriks (untuk *multivariate time-series*). Berikut adalah opsi penggunaan fungsi ***ts***:

```
ts(data = NA, start = 1, end = numeric(0), frequency = 1,
deltat = 1, ts.eps = getOption("ts.eps"), class, names)
```

Contoh:

```
> ts(1:10, start = 1959)
Time-series:
Start = 1959
End = 1968
Frequency = 1
[1] 1 2 3 4 5 6 7 8 9 10
> ts(1:47, frequency = 12, start = c(1959, 2))
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
1959 1 2 3 4 5 6 7 8 9 10 11
1960 12 13 14 15 16 17 18 19 20 21 22 23
1961 24 25 26 27 28 29 30 31 32 33 34 35
1962 36 37 38 39 40 41 4243 44 45 46 47
```

List

List adalah suatu vektor terurut dari sekumpulan komponen. Setiap komponen dapat berupa sembarang data objek, yaitu vektor, matriks, *data frame*, atau *data lists* endiri. Tiap komponen pada data *list* dapat mempunyai *length* dan mode yang berbeda. Membentuk *list* menggunakan cara yang sama dengan membuat *data frame*, hanya fungsi yang digunakan adalah ***list***.

Contoh: dari variabel *x* dan *y* yang didefinisikan pada contoh sebelumnya:

```
> L1 <- list(x, y); L2 <- list(A = x, B = y)
> L1
[[1]]
[1] 1 2 3 4
[[2]]
[1] 2 3 4
> L2
$A
[1] 1 2 3 4
```

$B
[1] 2 3 4
> names(L1)
NULL

> names(L2)
[1] "A" "B"

1.4.2 Membaca Data dalam File

Dalam membaca dan menulis *file*, *R* menggunakan *working directory*. *Working directory* adalah tempat di mana *R* membaca dan menyimpan *file* kerja kita secara *default*. Untuk mengetahui *working directory* yang sedang digunakan, gunakan perintah ***getwd()***. *Working directory* ini bisa diubah menggunakan perintah ***setwd(****path directory****)***, misal ***setwd(****"D:/data"****).***

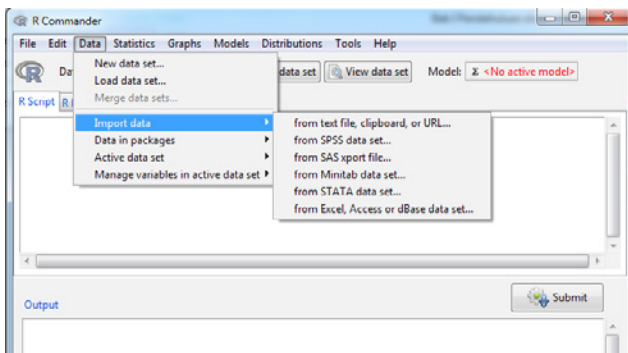
R bisa membaca data yang disimpan sebagai *file* teks (ASCII) menggunakan fungsi ***read.table***, ***scan***, dan ***read.fwf***. Fungsi ***read.table*** memiliki beberapa variasi, yaitu: ***read.csv***, dan ***read.delim***. Opsi penggunaan fungsi ***read.table*** diberikan sebagai berikut:

read.table(file, header = FALSE, sep = "", quote = "\"'",
dec = ".",
row.names, col.names, as.is = FALSE, na.strings = "NA",
colClasses = NA, nrows = -1,
skip = 0, check.names = TRUE, fill = !blank.lines.skip,
strip.white = FALSE, blank.lines.skip = TRUE,
comment.char = "#")

Jika fungsi ***read.table*** akan menghasilkan suatu *data frame*, sehingga tujuan utamanya adalah untuk membaca data dalam bentuk tabular. Adapun fungsi ***scan*** lebih fleksibel, kita bisa menentukan mode dari variabel-variabel yang kita impor. Opsi penggunaan fungsi ***read.table*** diberikan sebagai berikut:

```
scan(file = "", what = double(0), nmax = -1, n = -1, sep = "",
quote = if (sep=="\n") "" else "'\"", dec = ".",
skip = 0, nlines = 0, na.strings = "NA",
flush = FALSE, fill = FALSE, strip.white = FALSE, quiet =
FALSE,
blank.lines.skip = TRUE, multi.line = TRUE, comment.
char = "",
allowEscapes = TRUE)
```

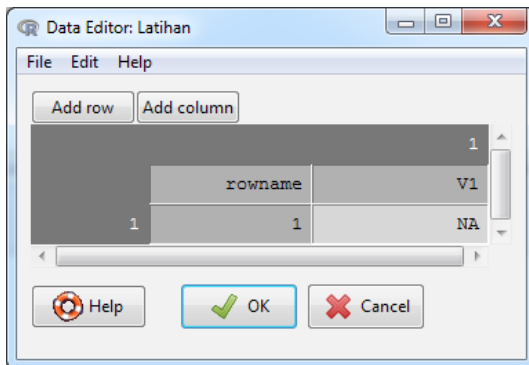
Sementara itu, fungsi ***read.fwf*** digunakan untuk membaca/mengimpor *file* dengan data yang memiliki format *fixed width*. Berikut adalah opsi penggunaannya:

```
read.fwf(file, widths, header = FALSE, sep = "\t",
as.is = FALSE, skip = 0, row.names, col.names,
n = -1, buffersize = 2000, ...)
```

Gbr 1.10 Menu impor data menggunakan R-commander

Selain membaca atau mengimpor *file* data menggunakan *command line R*, kita juga bisa mengimpor *file* data menggunakan *R-commander*. Program yang format datanya dapat dibaca oleh **R** adalah data dari *file* **teks** atau **clipboard**, dataset **SPSS**, dataset **MINITAB**, dataset **STATA**, data dari **Excel**, **Access**, atau **dBase**.

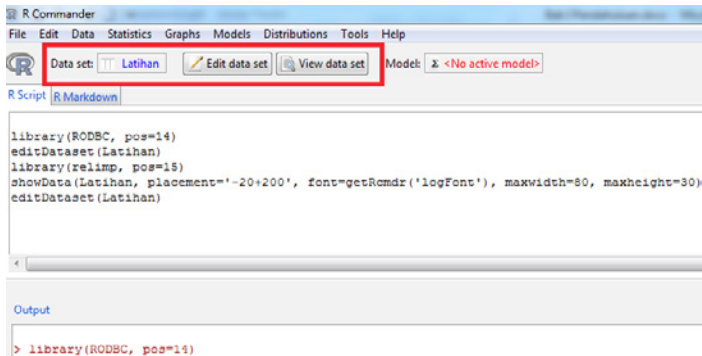
1.4.3 Input Data Menggunakan R-Commander

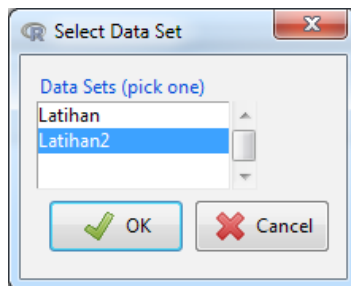
Selain mengimpor *file* data, dalam *R-commander* kita juga bisa menginputkan langsung data (entri data) yang akan digunakan. Pengisian data secara langsung di dalam **R** dengan menggunakan **R-commander** dapat dilakukan melalui menu **Data**, dan pilih **NewData set ...** Proses input data dilakukan melalui suatu jendela **Data Editor** yang akan terbuka, seperti terlihat pada Gambar 1.11.

Tampilan jendela Data Editor awalnya hanya terdiri dari 1 baris dan 1 kolom, namun kita bisa menambahkan jumlah baris dan kolom sesuai kebutuhan dengan mengklik tombol **Add row** atau **Add column**. Pada dasarnya, proses pengisian data ini adalah sama dengan Data Editor pada paket statistik yang lain.

Gambar 1.11 Tampilan Data Editor

1.4.4 Melihat, Mengedit, dan Memilih Data Set yang Sedang Aktif

Untuk mengetahui data set yang sedang aktif di memori, bisa dilihat pada jendela utama R di bawah menu (lihat bagian yang ditandai kotak merah pada Gambar 1.12). Untuk melihat atau menampilkan data set tersebut, klik tombol **View data**

set, sedangkan untuk mengedit data set maka klik tombol **Edit data set**. Pada *command line* kita bisa mengedit data set dengan menampilkan jendela data editor menggunakan perintah **edit**, sedangkan untuk menampilkan data set dengan menggunakan perintah **showData**.

Gambar 1.12 Data Set Aktif dan Tombol untuk Mengedit dan Menampilkan Data Set

Jika kita ingin bekerja dengan data set yang lain, maka kita harus mengganti data set yang aktif pada *R-commander*. Untuk memilih data set yang aktif, klik menu **Data**, arahkan kursor ke **Active data set**, kemudian klik **Select active data set...** . Selanjutnya, akan muncul jendela dialog untuk memilih data set, pilihlah data set yang dikehendaki.

Gambar 1.13 Jendela Dialog Memilih Data Set